Clustering Algorithm for Predicting Quantum States in a Quantum Computer

Undergraduate Research Thesis

by

Feng Qian

Advisor: Daniel J. Gauthier

the Ohio State University

April, 2021

Abstract

Quantum state tomography refers to the process of estimating the density matrix of an unknown system after obtaining data through a series of quantum measurements. Quantum state tomography can be divided into two processes: quantum measurement and reconstruction algorithm. From these two processes, different schemes can be designed, and they also have their own advantages and disadvantages. At the same time, quantum tomography is mainly concerned with two factors: estimation accuracy and complexity. High-precision quantum tomography is a necessary condition for quantum computing and other quantum technologies, and the scarcity of quantum resources forces us to find more effective methods to reduce estimation errors. The complexity problem of quantum tomography grows with the dimensionality of the quantum system.I simulate the behavior of a linear optics quantum computing circuit that performs the quantum Fourier transform. We investigated the possibility of using k-means clustering algorithm to predict the output of quantum computer and compare its performance with conditional generative adversarial network (CGAN), shadow tomography, maximum likelihood esitmation. Using this simulated data, I apply the various algorithms to the order-finding problem to determine the trade-off between required number of measurements and the accuracy. The result shows that k-means clustering can take a polynomial number of samples to access the result of order-finding problem with more than 90 percent accuracy. The comparsion of all four methods is analyzed.

Contents

List of Figures

Chapter 1

Introduction

Quantum computing is considered to be one of the most promising technologies in the 21st century, and the realization mostly involves processes such as the preparation, manipulation, and measurement of quantum states. In order to observe and test the actual effects of these quantum technologies, quantum state tomography have been established and have become basic tools in the field of quantum information [4].

Different from classical physics, quantum physics has the characteristics of quantum superposition, quantum entanglement, and no-cloning. Based on these basic physical principles, many new quantum technologies have been created, and most of them have advantages that classical information technology cannot match. For example, quantum secure communication that cannot be eavesdropped theoretically, quantum precision measurement that exceeds the limit of quantum noise in classical measurement, and quantum computer that solves some complex problems tens of millions of times faster than classical computers, etc [5] [6]., these technologies have all demonstrated great research value and application prospects.

The research on them also allows us to understand the quantum world more deeply and promotes the development and perfection of quantum physics. Similar to how we need to measure a classical system to perceive and use it, when researching and

applying a quantum system, we often want to know the relevant information or the state information of it.

The mathematical description of a quantum system is usually expressed in quantum states, which is one of the basic assumptions of quantum mechanics. Because it is very different from the description method of classical physics, people have developed a new branch of quantum precision measurement [7]. For example, for the measurement of quantum phase [8], the method using quantum properties can achieve higher accuracy than the classical method, and one of its important applications is gravitational wave detection [9]. Similarly, the realization of technologies such as quantum information and quantum computing relies on the precise preparation, operation, and measurement of quantum states.

In order to test the effects of real preparation in experiments, the concept of quantum state tomography was proposed [15,16]. It obtains the required information through a series of quantum measurements, and then estimates the quantum state prepared in the experiment based on this information. Therefore, quantum state tomography is widely used and studied as one of the indispensable basic tools of quantum science.

In classical physics, we can deterministically obtain system state information through a single measurement or perform multiple measurements without changing the system state. However, this is not the case in quantum physics. Due to the existence of quantum superposition characteristics, quantum systems may be in different states at the same time. According to the Copenhagen School's interpretation, what we obtain is the probability that the quantum system is in these states. At the same time, the measurement will cause irreversible damage to the quantum state, that is, the collapse of the wave function, which makes it impossible for us to repeatedly measure the system to obtain information, so the quantum state tomography task becomes very complicated. In fact, we need to prepare many identical systems in

the same quantum state, commonly known as quantum copies, and select a set of measurement bases that meet the conditions of complete information for projection measurement, and then estimate the unknown quantum state based on the statistical results obtained.

In response to this research, Fano proposed in 1957 a method of using a density matrix to represent quantum state information, which laid a basic framework and a solid theoretical foundation for the development of quantum tomography techniques [10]. With the deepening of later research and a large number of applications in experiments, quantum state tomography techniques has blossomed everywhere, and different technologies have been proposed one after another.

An earlier method is to use the Bonn rule to establish the linear relationship between the measured data and the unknown quantum state to solve the density matrix, that is, the linear reconstruction method. Its advantages are simple and fast, but there is often no physical meaning for the corresponding density matrix due to data errors.

In response to this problem, the maximum likelihood estimation is proposed [11], which ensures that the estimation results meet all physical requirements, and with its clear meaning and high accuracy, it is quickly applied in most experiments. But then Blume-Kohout found that the maximum likelihood estimation tends to give a density matrix with zero eigenvalues, which is unreasonable in some cases, for which he proposed a constrained maximum likelihood estimation.

Another difficulty faced by quantum state tomography is that the dimensionality of the Hilbert space increases exponentially with the number of qubits, that is, the exponential increase of the parameter to be estimated. This makes it necessary to select a large number of measurement bases in the experiment and the complexity of the reconstruction algorithm in the later data processing is also very high.

There is no general good solution to this problem, but when certain conditions are met or there is certain a priori information about the target quantum system, such as the pure state required by most quantum technologies, the symmetrical quantum state or the direct matrix product state. We can often create some new tomographic techniques to simplify the task [10] [12] [13]. For example, in recent years, the shadow tomography has received great attention. As long as the task doesn't require to reconstruct complete quantum state, shadow tomography can use few measurements to predict many properties such as expectation value of local observable [3].

This thesis will focus on investigating the possibility of using k-means clustering to predict to result of quantum computer. The performance of various reconstruction algorithms of quantum state tomography will be compared in linear optical systems by numerical simulation. The first chapter of the paper is an introduction to this article, and the second one will give a general introduction to the quantum computing and implementation of linear optical quantum computing platforms. In Chapter 3, k-means clustering and three modern quantum tomography (Shadow tomography, Neural network tomography, Maximum likelihood estimation) techniques will be introduce. In Chapter 4, we will discuss simulation methods and possible error models in quantum optical circuit. Chapter 5 will discuss their performance in predicting the results of the order finding problem.

Chapter 2

Introduction to Quantum Computing

This chapter is an introduction about basic concepts about quantum computing, quantum Fourier transform algorithm, Shor's Algorithm, and linear optical circuit implementation of quantum computing.

2.1 Quantum Bits

A quantum bit (qubit) is the fundamental building block of quantum computer. Not the same as the classical bit which can only stay in 0 or 1 state, the quantum bit can be in superposition of $|0\rangle$ and $|1\rangle$ state.

$$|\psi\rangle = \alpha\,|0\rangle + \beta|1\rangle \tag{2.1.1}$$

Where α and β are complex number with restriction:$|\alpha|^2 + |\beta|^2 = 1$. If one can have two qubits, then they will be able to construct a 4 states: 00, 01, 10, 00. The

two qubits system can be expressed as:

$$|\psi\rangle = \alpha\,|00\rangle + \beta\,|01\rangle + \gamma\,|10\rangle + \delta\,|11\rangle \tag{2.1.2}$$

In which $\alpha^2 + \beta^2 + \gamma^2 + \delta^2 = 1$ In general, for a quantum system with N qubits, there are 2^n states that can be used to encode information and perform computation spontaneously.

2.2 Quantum Fourier Transform and Shor's Algorithm

On a quantum computer, to decompose the integer N into prime factors, the operation of the Shor algorithm requires polynomial time. More precisely, this algorithm takes $O((logN))$ time [14], showing that the prime factorization problem can be solved in polynomial time using a quantum computer, so it is in the complexity class BQP. This is faster than the traditional known fastest factorization algorithm, the ordinary number field screening method is faster by an exponential difference.

The Shor's algorithm is very important because it represents the use of quantum computers, we can use it to crack the widely used public key encryption method, that is, the RSA encryption algorithm. The basis of the RSA algorithm is the assumption that we cannot decompose a known integer efficiently. As far as we know, this assumption is true for classical computers; there is no known traditional algorithm that can solve this problem in polynomial time.

However, Shor's algorithm shows that the problem of factorization can be solved efficiently on a quantum computer, so a sufficiently large quantum computer can crack RSA. This is a very big motivation for the establishment of quantum computers and the research of new quantum computer algorithms. Multiply two prime numbers,

such as 907*641=581387, and use a computer to process it. It seems that there is no difficulty. But if I give you 581387 and ask you to find its prime factor, the problem becomes very complicated. Maybe you can use a computer to try them one by one, but when the numbers become bigger, reaching hundreds or thousands, even computers can't do anything.

There are many questions in the world like this, it is difficult to find the answer, but once the answer is found, it is easy to verify. Similar problems are called NP problems. The NP problem is difficult to deal with because its time complexity is often exponential. This means that as the scale of the problem increases linearly, the time required increases exponentially. Using this principle, people created the RSA algorithm, which uses the principle that large numbers are difficult to decompose but easy to verify to effectively encrypt data. Quantum computers have the ability to reduce the time complexity of "prime factor decomposition" to the polynomial level, making it possible to solve the problem of large number decomposition. This is Shor's algorithm. The proposal of the Shor algorithm means that the security of RSA keys has been challenged. Below we will introduce the content of Shor's algorithm [14].

Shor's algorithm first converts the prime factor decomposition problem into a sub-problem. Let's look at the conversion process of the problem. Suppose the number we want to decompose is N:

STEP 1: Take a random positive integer $1 < a < N$ and define a function: $f(x) = a^x mod N$;

STEP 2: This function must be a periodic function, and find that its period is r. (It's called order finding problem);

STEP 3: If r is odd, then go back to STEP 1. If r is even, then calculate $f(r/2)$;

STEP 4: If $f(r/2) = -1$, then go back to STEP 1. Otherwise, calculate the greatest common divisor of $f(r/2) + 1$ and $f(r/2) - 1$respectively for N;

STEP 5: These two greatest common divisors are the two prime factors of N;

For explaining the step, I will break out it into a few steps to explain. The first is a set of Hadamard transforms, they only act on a set of N working bits, so this total state will become

$$|\text{Working}\rangle|\text{Ancilla}\rangle = \left(\sum_{x=0}^{2^N-1} |x\rangle\right) |00...001\rangle$$

When a quantum function acts on this set of quantum states, it is equivalent to that all the values of the independent variable of this function from 0 to $2^N - 1$ are saved to the auxiliary bits. We know that $f(x) = a^x mod N$ is a periodic function, assuming its period is T. Obviously, $f(x) = f(x+T) = f(x+2T)....|x\rangle|f(x)\rangle + |x+T\rangle|f(x+T)\rangle + |x+2T\rangle|f(x+2T)\rangle + .. = (|x\rangle + |x+T\rangle + |x+2T\rangle + ...)\,|f(x)\rangle$
Back to $a = 7, N = 15$ case, $|\text{Working}\rangle|\text{Ancilla}\rangle = (|0\rangle + |4\rangle + |8\rangle + ...)|1\rangle + (|1\rangle + |5\rangle + |9\rangle + ...)|7\rangle + (|2\rangle + |6\rangle + |10\rangle + ...)|4\rangle + (|3\rangle + |7\rangle + |11\rangle + ...)|13\rangle$

Because this state is an entangled state, when we measure the auxiliary bit, the working bit will collapse to the corresponding situation. But no matter what the measured value of the auxiliary bit is, the working bit will always be reserved as a superposition state where each component is exactly a set of cycles. Then the period of the number represented by this group of superposition states will be quickly completed by the quantum Fourier transform.

The period of the state can be quickly completed by the quantum Fourier transform. Let's take $|0\rangle + |4\rangle + |8\rangle + ...$ as an example to see how the quantum Fourier transform is done, and then you will find that it is for 1,5,9,13... Or 2, 6, 10, 14... can get similar results.The quantum Fourier transform has two important parts, the first is the recursive sequential control rotation (CROT) operation, and the second part is to change the order of the bits. In mathematical expression, each item is processed in the form of discrete Fourier transform. $y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j \omega^{jk}$ Where x_j represents the j-th component of the input quantum state, and k represents the component of

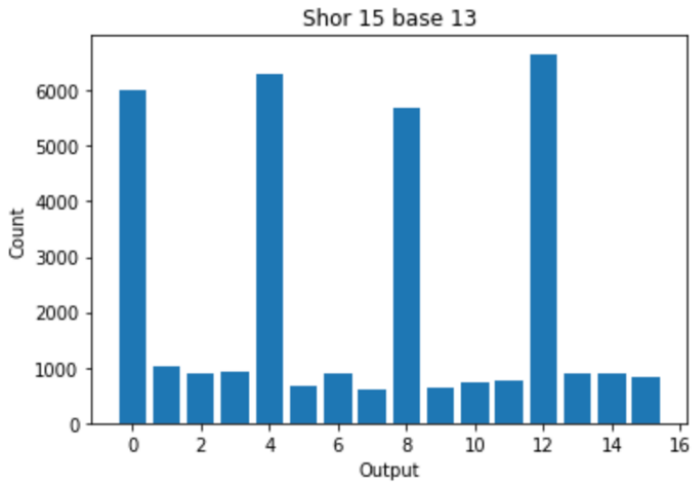

Figure 2.1: Example of output from order finding problem in factoring 15 with base 13. The vertical axis is the number of count for each state. The horizontal axis is decimal expression for each fock state, for example, $|0010\rangle \implies |2\rangle$

the output quantum state. If it is represented by n qubits, then $\omega = e^{\frac{2\pi i}{2^n}} = e^{\frac{2\pi i}{N}}$. From the matrix point of view, it is

$$F_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(N-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(N-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \omega^{3(N-1)} & \cdots & \omega^{(N-1)(N-1)} \end{bmatrix} \tag{2.2.1}$$

Assume that there are only 4 working bits. Then the input quantum state is $|\text{Input}\rangle = |0\rangle + |4\rangle + |8\rangle + |12\rangle$, in which the 0,4,8,12 are the decimal representation of the binary state representation. This means that $x_0 = x_4 = x_8 = x_{12} = 1$, and $\omega = e^{2\pi i/16}$, all

other components are 0. According to the Fourier transform formula, we can calculate

$$\begin{aligned} y_k &= \frac{1}{\sqrt{4}}(\omega^{0k} + \omega^{4k} + \omega^{8k} + \omega^{12k}) \\ &= \frac{1}{2}(1 + i^k + (-1)^k + (-i)^k) \end{aligned} \tag{2.2.2}$$

Here is the value of each component (the kth component) in the output state after the quantum Fourier transform of the working bit is performed. So we can get $y_0 = y_4 = y_8 = y_{12} = \frac{1}{2}$, in other cases $y_k = 0$ $(k \neq 0, 4, 8, 12)$, then the final output quantum state Then $|\text{Output}\rangle = |0\rangle + |4\rangle + |8\rangle + |12\rangle$ We can use continued fraction decomposition to get period. In the final measurement, we will randomly get one of the four results 0, 4, 8, 12, but this result is not a period. But the result of the quantum Fourier transform reveals one thing: $\omega^{irx} = e^{2\pi irx/2^N} \sim 1$.

Among them, we assume that the measurement result is x, the total number of working bits is N, and the period of the function is r. Then we have $x/2^N = c/r$ Where c is an unknown integer. So we can find the function period approximately by this formula. For example, $x = 4$, $N = 4$, we have $c/r = 1/4$ In this way, we have found the period $r = 4$. So far, the quantum computer part of Shor's algorithm has been solved. You can check whether the period of the function $f(x) = 7^x(mod15)$ is indeed 4. You can also check whether the greatest common factor of $f(r/2) + 1$ and $f(r/2) - 1$ and 15 is the prime factor of 15. Sometimes $x/2^N$ may not be able to smoothly approximate a reasonable r, so we can get an approximate score through the continued fraction decomposition method to obtain r.

2.3 Linear Optical System

The linear optical implementation of the qubit is by using dual rail construction and unitary transformations are using Beam splitter (BS) and Phase shifter (PS)

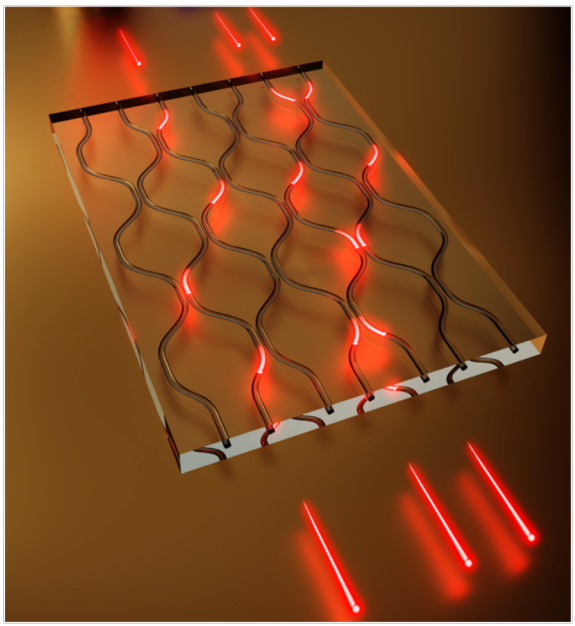

Figure 2.2: Photonic Circuit represneting real world experiment. The redline represents photons, the tube represents the waveguide, and the joint between each tube is beam splitter and beam shifter. [1]

[14].The $|0\rangle$state and $|1\rangle$ state are represented by the two input and two output ports beam splitters. A photon is injected in input port $\hat{\mathbf{a}}_1$ of the BS represented the state $|0\rangle$ and a photon is injected in input port $\hat{\mathbf{a}}_2$ of the the BS represented the state $|1\rangle$. In Fig. 2.2, every two input/output port represent 1 qubit. We can only inject one photon for each two input port. Each different photon series represent a different quantum state. The beam splitter and phase shifter are connected by waveguide. In Fig. 2.2, the waveguide is the Every time we want to execute a quantum algorithm, we need to inject photon in fixed configuration many times and measure which output port has photon comes out. Then, we can know what the outcome is.

Previously, it has been shown that multiport beam splitter (BS) and phase shifter(PS) configuration can be used to perform any unitary transformation in Hilbert space of any finite dimension [15]. The 50:50 BS and a PS will construct a Hadamard gate

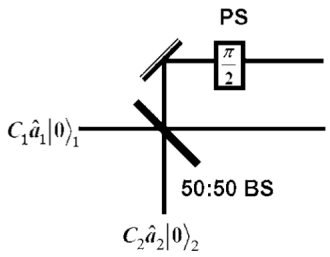

Figure 2.3: Photonic Circuit represneting a qubit going through a Hadamard gate [2]

which is one the most elementary gate in quantum computing.

$$\begin{pmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \end{pmatrix} = \exp(-i\alpha) \begin{bmatrix} \cos(\beta)\exp(i(\gamma+\delta)) & \sin(\beta)\exp(i(-\gamma+\delta)) \\ -\sin(\beta)\exp(i(\gamma-\delta)) & \cos(\beta)\exp(-i(\gamma+\delta)) \end{bmatrix} \begin{pmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{pmatrix} \quad (2.3.1)$$

Any two qubits gate can be realized by using one BS and one PS. The BS can be treated as a unitary matrix U (0, β, γ, δ) in which any incident beam goes through a phase shift γ, then the amplitudes are rotated by β, and finally the phases are shifted again by an angle δ. The α is the transformation made by the phase shifter [2].

In order to implement quantum Fourier transform in linear optics circuit, we need to use a Cooley-Tukey algorithm for the design of optical circuit. The Cooley-Tukey algorithm can re-express the Fourier transform of an arbitrary composite size $QFT(n) = S_n^{(n)} S_{n-1}^{(n)}, ..., S_2^{(n)} S_1^{(n)}$. The $S_n^{(n)}$ is the n-th decomposed matrix of the original matrix QFT(n).

The way to design the optical circuit is

Step 1: Calculate the quantum Fourier transform matrix from equation (2.2.1).

Step 2: Break down the QFT matrix into n matrices, n is the number of qubits. Each matrix represent 2^{n-1} BS-PS components

Step 3: Using the Cooley-Tukey algorithm to find the graphical representation of the optical circuit.

Step 4: Calculate the parameter of each BS-PS component based on $S_k^{(n)}$

Here is a example of 2 qubits case:

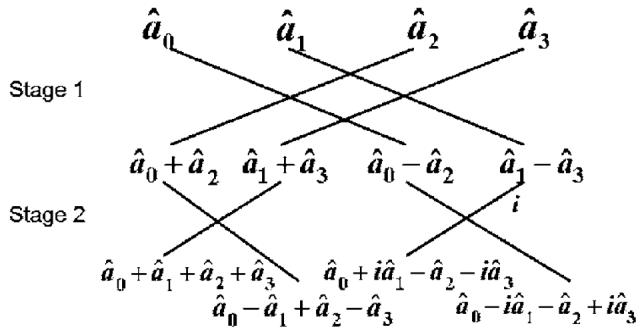

Figure 2.4: The graphical representation of the QFT using BS-PS components of input operator to get to output operator in each stage Figure taken from [2].

1. Based on equation (2.2.1), the QFT matrix is

$$QFT(2) = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$$

2. Then, we can find the decomposed matrix $S_1^{(2)}$ and $S_2^{(2)}$ based on Cooley-Tukey algorithm.

$$S_1^2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}; S_2^2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & i \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -i \end{bmatrix}$$

3. The graphical representation is fig 2. 4. Three of the four BS–PS components are exactly the same as those of the 2D DFT,but the fourth BS–PS component is different in the sense that a different PS has been introduced (the BS is still balanced) with half the phase shift compared to the others:

$$BS_2^2 = U(\frac{\pi}{4}, \frac{\pi}{4}, \frac{7\pi}{4}, \frac{\pi}{2})$$

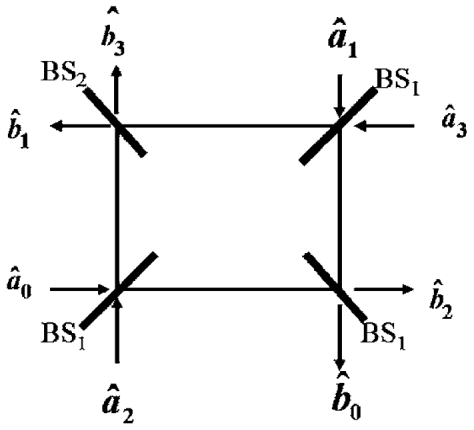

Figure 2.5: The optical circuit design of 4d QFT matrix in which $\hat{\mathbf{a}}_i$ is the input operator, $\hat{\mathbf{b}}_i$ is the output operator. Figure taken from[2].

In this notation $BS_1 == U(\frac{\pi}{2}, \frac{\pi}{4}, 2\pi, \frac{\pi}{2})$where the upper index represents the number of qubits in the DFT, and the lower index represents the different BS–PS components.

Chapter 3

Quantum Tomography

The quantum computer is a probabilistic machine. In order to get access to the result, we need to measure the output wave function in different state. With the increasing size of the system, the required number of the measurements will increase exponentially. There is way to reduce the total number of measurements that need to be performed which is Quantum tomography. The quantum state tomography is made by two part: measurement and reconstruction algorithm. In this paper, we will mainly introduce the reconstruction algorithm. We know from quantum mechanics that the state of a quantum system is generally described by a wave function, which can be written in the form of a state vector using Dirac notation. Observing a certain physical quantity of the wave function will often produce multiple different results, which can be expressed as the following form:

$$|\psi\rangle = \sum_{i=1}^{d} c_i \, |\phi_i\rangle$$

The modulus square $|c_i|^2 = |\langle\psi|\phi_i\rangle|^2$ of its coefficient represents the probability that the quantum state will collapse to $|\phi_i\rangle$ after measurement. The above formula shows the quantum superposition which is different from classical physics, and the same quantum state can also be expressed in different forms under different observation

bases. If the quantum state corresponding to these measurement results can expand the Hilbert space where the quantum system is located, after using $|\phi_i\rangle$ as the base, then the quantum state $|\psi\rangle$ can be written as $(c_1, c_2, ..., c_d)^T$ under the base, which corresponds to one of vector in the Hilbert space, this concise and clear representation can bring us convenience in many cases. The above description method can only be used when the quantum system is in pure state. But for a quantum system, the quantum state of different particles may be different. For such a system, we call it a mixed state. The mixed state is represented by the density matrix given by

$$\rho = \sum_{j=1} p_j \left|\phi_j\right\rangle \left\langle\phi_j\right|$$

Among them, p_j represents the probability that the particle in the quantum system is in the quantum state $|\phi_j\rangle$, and $\sum_j p_j = 1$. Similarly, if we select a complete set of quantum states as the basis of the Hilbert space and write it in the form of a vector, then ρ will be expressed as a square matrix but for a probabilistic measurement, we can only get the accurate the probability for each by increasing the number of measurements. By using the reconstruction algorithm, we can only use fewer samples to get similar probability distribution. Similarly, the pure state can also be written in the form of a density matrix, so this provides a convenient and universal way of describing quantum systems. But not any density matrix can correspond to a physically existing quantum state. It can be proved that a density matrix has a physical correspondence if and only if it satisfies the following three conditions:

a. Hermitian $\rho^\dagger = \rho$:;

b. Normalization: $\mathrm{Tr}(\rho) = 1$;

c. Semi-positive definite: $\rho \geq 0$ that is, all eigenvalues of the density matrix are non-negative. As long as we know the density matrix, the probability density of each state will be the element on the diagonal.

3.1 Maximum Likelihood Estimation

In the past, people use linear regression method in Quantum tomography. However, because the measurement data inevitably has certain errors, the reconstructed density matrix often has an eigenvalue less than zero, which violates the semi-definite requirement of the density matrix. For example, the pure state on the Bloch sphere of a single qubit may cause the reconstruction result to fall outside the Bloch sphere due to a slight disturbance of the measurement error, that is, the corresponding density matrix contains eigenvalues less than zero, which makes the density of reconstruction The problem that the matrix exceeds the allowable physical space of the quantum state is called the non-physical problem of quantum state tomography. To solve this problem, scientists introduced the maximum likelihood method in statistics [11]. First, we can write the unknown density matrix as:

$$\rho = T^{\dagger}T/\mathrm{Tr}\big(T^{\dagger}T\big) \tag{3.1.1}$$

It can be proved that for any square matrix T, ρ satisfies the three requirements of the density matrix in the previous section, namely Hermitian, normalization, and semi-definiteness. Next, the square matrix T needs to be parameterized so that ρ in the Eq. (3.1.1) can express any quantum state. Taking a single bit as an example, it is generally expressed as an upper triangle or a lower triangle for computational convenience:

$$T(t) = \begin{bmatrix} t_1 & 0 \\ t_3 + it_4 & t_2 \end{bmatrix} \tag{3.1.2}$$

So ρ is expressed as a function of four real parameters of (t_1, t_2, t_3, t_4). It needs to be explained that although there are only three single-bit free parameters, the choice of four parameters here is to ensure that this parameterization method can make ρ express any single-bit density matrix, and through the normalized transformation of

formula 3.1.1, The actual free parameters of ρ are still three. Similarly, for the density matrix of the d-dimensional Hilbert space, we can express it in the upper or lower triangular form of d^2 real parameters. After satisfying the three requirements of the density matrix, the next step is to solve each parameter based on the measured data. We introduce the likelihood function $\mathcal{L}(t_1, t_2, t_3, t_4)$, which represents the probability of observation when the quantum state is $\rho(t_1, t_2, t_3, t_4)$

Assuming that the noise in the experiment satisfies the Gaussian distribution, then:

$$\mathcal{L}(t_!, t_2, t_3, t_4) = \Pr(n_i|\rho) = \frac{1}{A}\prod_i \exp\left[-\frac{(n_i - \bar{n_i})^2}{2\sigma_i^2}\right]$$

Among them, A is the normalization constant, n_i represents the i-th measurement data, and, $\bar{n_i} = N\langle\phi_i|\rho|\phi_i\rangle$ represents the expected value of the i-th measurement data, σ_i is the standard deviation of the corresponding Gaussian distribution, and generally takes $\sigma_i \approx \sqrt{\bar{n_i}}$. According to the idea of the maximum likelihood method, the value of the parameter to be estimated should maximize the value of the likelihood function, so we can express the quantum state estimation as the following problem of finding the extreme value:

$$t_1, t_2, t_3, t_4 = argmax\frac{1}{A}\prod_i \exp\left[-\frac{(n_i - N\langle\phi_i|\rho(t_1, t_2, t_3, t_4)|\phi_i\rangle)^2}{2N\langle\phi_i|\rho(t_1, t_2, t_3, t_4)|\phi_i\rangle}\right]$$

After we know the t_1, t_2, t_3, t_4, we can calculate the matrix T(t) and ρ. Therefore, we can get the estimation of the density matrix of the quantum state.

3.2 Classical Shadowing Tomography

For the n-particle system $D = 2^n$, then the density matrix ρ will be a $2^n * 2^n$matrix, so the number of measurements we need to reconstruct ρ is $\Omega(D^2)$, which is a necessary condition. On STOC 2016, O'Donnell-Wright [16] and Haah- Harrow-Ji-Wu-Yu

Algorithm 1 *Median of means prediction* based on a classical shadow $\mathsf{S}(\rho, N)$.

function LINEARPREDICTIONS($O_1, \ldots, O_M, \mathsf{S}(\rho; N), K$)
  Import $\mathsf{S}(\rho; N) = [\hat{\rho}_1, \ldots, \hat{\rho}_N]$ ▷ Load classical shadow
  Split the shadow into K equally-sized parts and set ▷ Construct K estimators of ρ

$$\hat{\rho}_{(k)} = \frac{1}{\lfloor N/K \rfloor} \sum_{i=(k-1)\lfloor N/K \rfloor + 1}^{k\lfloor N/K \rfloor} \hat{\rho}_i$$

  for $i = 1$ to M **do**
    Output $\hat{o}_i(N, K) = \text{median}\left\{\text{tr}\left(O_i\hat{\rho}_{(1)}\right), \ldots, \text{tr}\left(O_i\hat{\rho}_{(K)}\right)\right\}$. ▷ Median of means estimation

Figure 3.1: The procedure of median of means in prediction based on classical shadow [3]

[17]respectively proved that this is also a sufficient condition.If the reconstructed density matrix is $\rho^{'}$, then their error can be described by the trace $||\rho - \rho^{'}|| \leq \epsilon$ [18]. gives a method, so that we only need $O((D/\epsilon^2)\ln(D/\epsilon))$ copies. But it is still an exponential number of copies. But what if only part of the information of the quantum state is reconstructed? The shadow tomography mentioned in Aaronson's [18] is an example. It is not the entire matrix to be reconstructed, but the measured results of each matrix in the given positive operator-valued measure (POVM); Because only a part of the density matrix is involved, it is called "shadow". In this work, It is proven that for a $D = 2^n$ dimension system, it only needs $O(\frac{\log 1/\epsilon}{\epsilon^5} \log^4 M \log D)$. To create a shadow tomography, one should apply a random unitary transformation on the state $\rho \implies U\rho U^\dagger$, and perform a computational basis measurement.Then, we can store the result of measurement and the information about the transformation U and the basis $\left|\hat{\mathbf{b}}\right\rangle$. The average mapping from ρ to $U^\dagger \left|\hat{\mathbf{b}}\right\rangle \left\langle\hat{\mathbf{b}}\right| U$ can be treated as a quantum channel:

$$E[U^\dagger \left|\hat{\mathbf{b}}\right\rangle \left\langle\hat{\mathbf{b}}\right| U] = M(\rho) \implies \rho = E[M^{-1}(U^\dagger \left|\hat{\mathbf{b}}\right\rangle \left\langle\hat{\mathbf{b}}\right| U)]$$

Repeating this process N times, the set of quantum channel relation will be a classical shadow of ρ:

$$S(\rho; N) = \{\rho_1 = M^{-1}(U_1^\dagger \left|\hat{\mathbf{b_1}}\right\rangle \left\langle\hat{\mathbf{b_1}}\right| U_1), ..., \rho_N = M^{-1}(U_N^\dagger \left|\hat{\mathbf{b_N}}\right\rangle \left\langle\hat{\mathbf{b_N}}\right| U_N)\}$$

Then, We use the median of means prediction based on classical shadow as comparison in this work. The procedure of this algorithm is in Algorithm 1 [3].

Different from two methods mentioned above to directly reconstruct the matrix, in the following section, I will introduce two method to indirectly get the result of the quantum Fourier transform. As we mentioned before, the output of quantum Fourier transform is a sparse distribution, and it is periodic. The goal for the quantum Fourier transform is to distinguish the period of each output result. Therefore, classification algorithm will also be one of the possible solution. In the following section, I will introduce two classification algorithm to predict the result of quantum Fourier transform.

3.3 Neural Network Tomography

In our comparison. we specifically discuss conditional generative adversarial network(CGAN). In [1] and [19], it has been shown that CGAN is able to adapt the noise and reconstruct underlaying state using up to two orders of magnitude fewer iterative step. Standard GAN framework is made by two neural network, one standard feed-forward neural network, and one standard classification neural network. The feedforward network with parameter θ_G, generates new data using noise vector z:

$$x^{'} = G(z; \theta_G)$$

The network is trained by the classify network with parameter θ_D which works as a discriminator to adjust the similarity between original data and generated data. The parameters θ_G, θ_D are optimized alternatively until the classify network can not distinguish the difference between generated data and real data. In each step, the θ_D

is optimized to reach its maximum at first.

$$E_{z\sim p_z}[\log(D(x;\theta_D))] + E_{z\sim p_z}[\log(1 - D(G(z;\theta_G);\theta_D))]$$

Then,θ_G is optimized to its minimum.

$$E_{z\sim p_z}[\log(1 - D(G(z;\theta_G);\theta_D))]$$

A traditional GAN can only generate data randomly but we can set up a condition variable c to improve the generated data. It has been shown the advantage in many previous works [20][19].

By using part of the simulated data, we can train the classifier to distinguish the period of the output. Then, we apply the classifier to the unseen simulated data to see the performance of this algorithm.

3.4 K-means Clustering Algorithm

The k-means clustering algorithm is an iterative solution clustering analysis algorithm. Its steps are to divide the data into K groups in advance, then randomly select K objects as the initial clustering centers, and then calculate The distance between each object and each seed cluster center, and each object is assigned to the cluster center closest to it. The cluster centers and the objects assigned to them represent a cluster. Each time a sample is allocated, the cluster center of the cluster will be recalculated based on the existing objects in the cluster. This process will be repeated until a certain termination condition is met. The termination condition can be that no (or minimum number) of objects are reassigned to different clusters, no (or minimum number) of cluster centers change again, and the sum of squared errors is locally minimum.

K-means Clustering Algorithm Procedure(example for 5*5 two photons)

1. Select the number of clusters, for example 3

2. Random select three (depend on the number of clusters you choose) start states$|0,0,0,1,1\rangle, |0,1,0,0,1\rangle, |1,1,0,0,0\rangle$

3. Find the minimum "distance" between the start point and each state one by one. For example, First, we find the distance between $|0,0,0,1,1\rangle$ and other states because there is no other cluster to compare so all states will belong to this cluster.

Second, find the distance between $|0,1,0,0,1\rangle$ (start point of the second cluster) and the other state. Then, some of state will go to the second cluster. Third, find the distance between $|1,1,0,0,0\rangle$ and the other states.

4. After completing above step, there will have three complete clusters. Then we need to "recenter" the cluster with point that is the mean of all state in the cluster. Doing this step many times until the center of the cluster no long changes. (Write a loop for tracking the difference for each iteration. End the loop until the difference is zero)

5. Finish the k-means clustering

6. This kind of method should only find the local minimum, so we need to some other method to find the global minimum. The one I use is to change the initial value of this process odd times. Then, we can find the global minimum.

The distance in here is defined as the

$$d = L_1 = \sum_{i=1}^{M} |\Psi_i - \Phi_i|$$

The i means the i the mode. ψ_i and Φ_i mean the occupation numbers of these two states in the i th mode. For example, $\psi = |1,1,0,0,0\rangle\, \Phi = |1,0,1,0,0\rangle$ The L_1 will be $(1-1)+(1-0)+|0-1|= 2$

By applying the output result of quantum Fourier transform, the data will be clustered. The center for each cluster will be near the high peak state. Then, we only need to calculate the shortest distance between two cluster to get the period.

Chapter 4

Monte Carlo Simulation

We will compare the performance of each quantum tomography method in the order-finding problem which is the quantum part of Shor's algorithm. In this chapter, the Monte Carlos method for order finding simulation in optical circuit will be introduced. Noise is an inevitable factor in experiment. In order to make model robust to those noise, some common noise model in optical circuit will be introduced and used in simulation as a part of comparison.

4.1 Monte Carlo Simulation

To simulate order finding problem, we assume that by using BS-PS components to build n-photon state (input state) given by

$$|\phi\rangle = \frac{1}{s^{\frac{1}{2}}} \sum_{j=0}^{s-1} \frac{(\hat{\mathbf{a}}_j^\dagger)^n}{\sqrt{n!}} \exp(if_i) |0\rangle$$

where n is the number photon, f_i is relative to $F^j(modN)$ by

$$f_i = 2\pi \frac{F_j(modN)}{N}$$

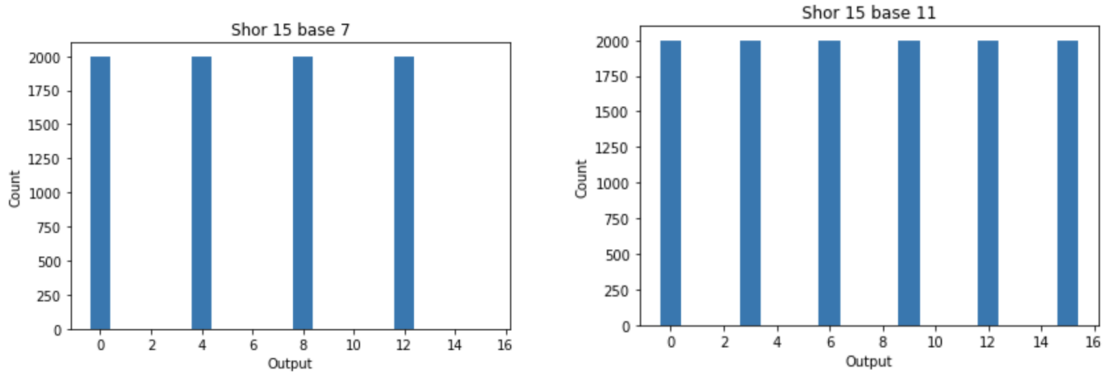

Figure 4.1: Noise free simulation results of order finding problem with based 7 and 11. The gap between each peaks is the period of the result.

in which N is the number to be factored. F is a number between 1 and N. In order to find the period, we need to apply quantum Fourier transform on $\hat{\mathbf{a}}_j^\dagger$. They are transformed to operator $\hat{\mathbf{b}}_k^\dagger$ as

$$\hat{\mathbf{a}}_j^\dagger \implies \frac{1}{\sqrt{s}} \sum_{k=0}^{s-1} \exp(2\pi jk/s) \hat{\mathbf{b}}_k^\dagger$$

then, we can get the output results from the following formula:

$$\left|\psi'\right\rangle = \sum_{j=0}^{s-1} \sum_{k=0}^{s-1} \exp(2\pi ijk/s) \hat{\mathbf{b}}_k^\dagger \exp(if_j) \left|0\right\rangle$$

According to this formula, we can calculate the probability for each state and store the probability in a list. Then, I use the function numpy.random.choices to generate random output state based on the probability we give in the list.

4.2 Error Model in Optical Circuit

In this section, we will introduce two kinds of noise model: thermal or environmental noise and photon loss.

In the real experiment, the quantum system will barely be a close system, there will always be thermal or experiment noise. We model this kind of noise by:

$$\rho_{mixed} = (1 - \sigma)\rho + \sigma\rho_{environment}$$

with $\sigma \in [0, 0.5]$. For simulating this case, I generate random unitary matrix $\rho_{environment}$ and add it to the QFT matrix. After that, I calculate the probability, generate random quantum state, and count the number of each state. The choice of σ is chosen randomly. The finial result is take average performance

For optical experiment, if the photon is made by a lossy resonator, the photon will be possible lose from the resonator before the measurement. We model this kind of noise based on previous work [1]by letting the original state evolved by time τ following the master equation:

$$\dot{\rho} = -\frac{i}{\hbar}[H, \rho] + \gamma\mathcal{L}[a]\rho$$

where $H = \hbar a^\dagger a$ is the free Hamiltonian,ω is the resonator frequency, γ is the photon loss rate, and $\mathcal{L}[a]\rho = a\rho a^\dagger - \frac{1}{2}a^\dagger a\rho - \frac{1}{2}\rho a^\dagger a$.

The formal solution for a density matrix from time t to $t + \tau$ is

$$\rho(t + \tau) = V_\tau(\rho(t))$$

where L is the Lindbald operators and

$$V_\tau() = exp(\tau L()),$$

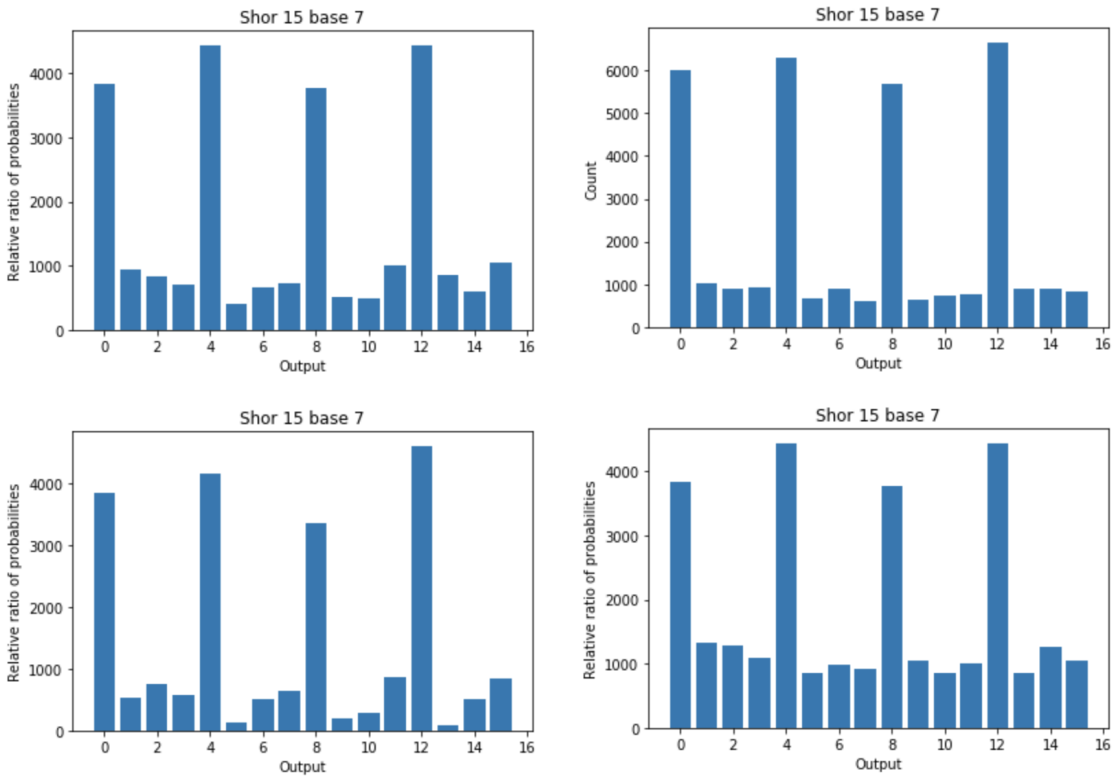

Figure 4.2: This four figures are including the noise model I mentioned above. The top left one has environment noise with factor $\sigma = 0.3$. The top right one has environment noise $\sigma = 0.2$. The bottom left one considered photon loss. The bottom right has both photon loss and environment noise$\sigma = 0.2$

We input operator in () to generate new operator. By taking first order in τ:

$$V_r = 1 + \tau \sum_{i=1}^{m} L_i()$$

Then, the density matrix with time evolution will be

$$\rho(T) = U V_\tau^i \rho(0) U^\dagger$$

According to this new error-included density matrix, we can generate the data with two kinds of model as mentioned in this chapter.

We can see that after we introduced the two kinds of error model in the simulation, it is not easy to tell the period when the total number of sample is small. Therefore, the performance under error-included data is important result to compare.

Chapter 5

Results

The k-means clustering algorithm has been shown that it can classify known error model in boson sampling validation [21]. The periodicity and sparsity are two main feature of order finding problem. We adopt k-means clustering, median of means prediction and CGAN to order finding problem by taking each period as different. Then,the algorithm will be performed to classify each different set of data in to different class.

In the simulation, we considered environment noise and photon loss. For each simulation, there are 30000 samples generated with random initial F. The size of the system is from 5 to 14 qubits. For the order finding problem, we choose to factor 15 with F = 7. According to simulation result, the initial guess F will reach maximum predict rate on the reconstructing the output result for all of the four techniques. Because the result of period finding will have sparse output result which will be a prior knowledge in our experiment. In Figure 5.4, it shows the comparison among four kinds of techniques. The vertical axis is total number count for each state. The horizontal axis is the decimal representation of each quantum state. Among them, the k-means clustering algorithm , CGAN and classical shadow tomography only require

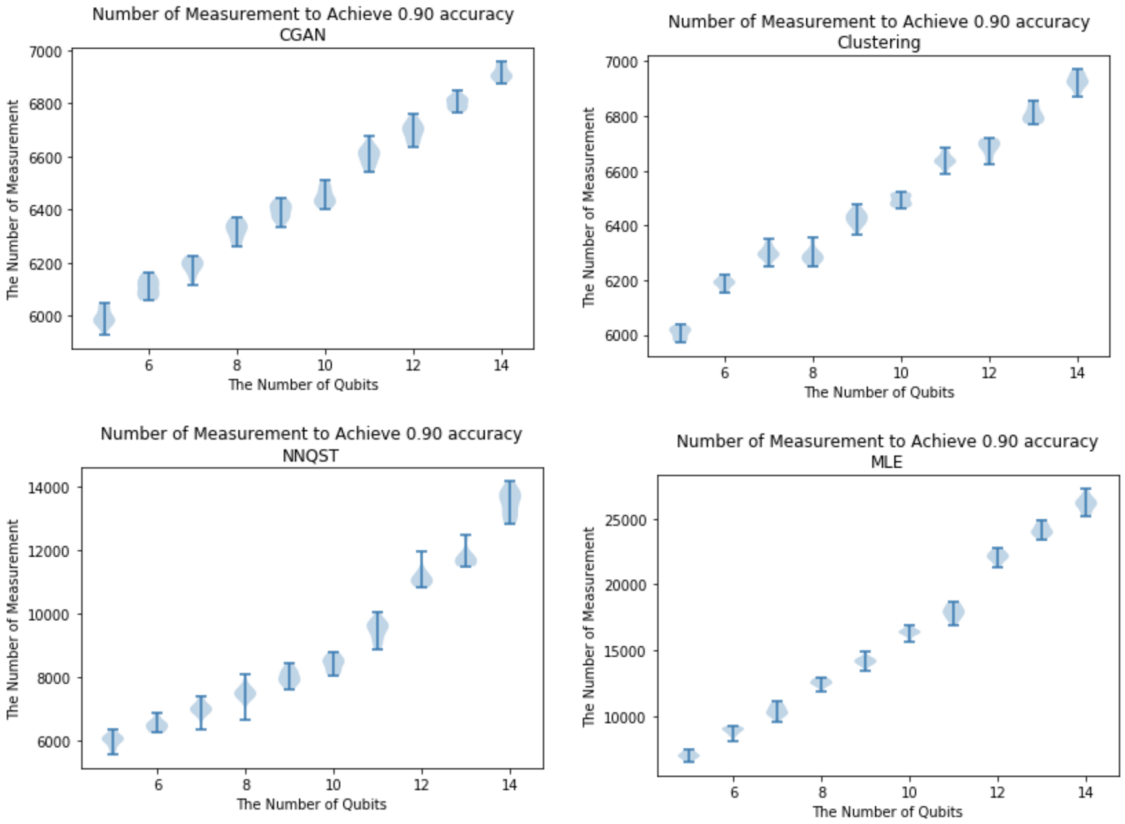

Figure 5.1: Violin plot for K-means Clustering algorithm, Maximum Likelihood, median of means prediction based on classical shadow, and CGAN

F/Method	K-means	CGAN	MLE	Classical Shadow
2	0.9879	0.9845	0.9453	0.9867
3	0.9851	0.9886	0.9446	0.9864
4	0.9821	0.9825	0.9413	0.9844
5	0.9824	0.9816	0.9444	0.9846
6	0.9871	0.9815	0.9411	0.9891
7	0.9921	0.9915	0.9513	0.9964
8	0.9861	0.9896	0.9463	0.9849
9	0.9841	0.9834	0.9401	0.9866
10	0.9826	0.9875	0.9484	0.9844
11	0.9247	0.9816	0.9393	0.9805
12	0.9881	0.9847	0.9469	0.9848
13	0.9867	0.9842	0.9452	0.9826
14	0.9871	0.9823	0.9423	0.9862

Table 5.1: The accuracy of each method with system size N = 10 in which F is the number we picked from zero to the number we need to factor.

polynomial samples to predict the result of order finding problem with the increasing number of qubits. To achieve 0.90 accuracy of classification, the Maximum likelihood method and neural network method need polynomial amount of samples and MLE has larger slope. For the CGAN and clustering, the required number of samples does not have obvious change. To achieve 0.99 accuracy, the required number of samples for MLE increase exponentially. The classical shadow tomography method takes only about 7500 examples and has no obvious change with increasing size of system. Compared to the CGAN, the increasing rate of clustering is smaller. However, for the CGAN, the training time the network is a disadvantage of it. In Fig. 5.2, the vertical axis is the size of the system, and the horizental axis is the hour it takes in google cloud computing services with 6 set of CPU and 12 set of GPU. It shows that CGAN take more than 4 hours to train the network with system size N = 10. Also, it will need to retrain the model if the size of the system size changes. In the semilog plot Fig. 5.4, we compare the runtime for each algorithm to achieve 0.9 accuracy. The prior knowledge about sparse output is not quantified which will play an important role in reducing the number of required samples. Compared with median of means

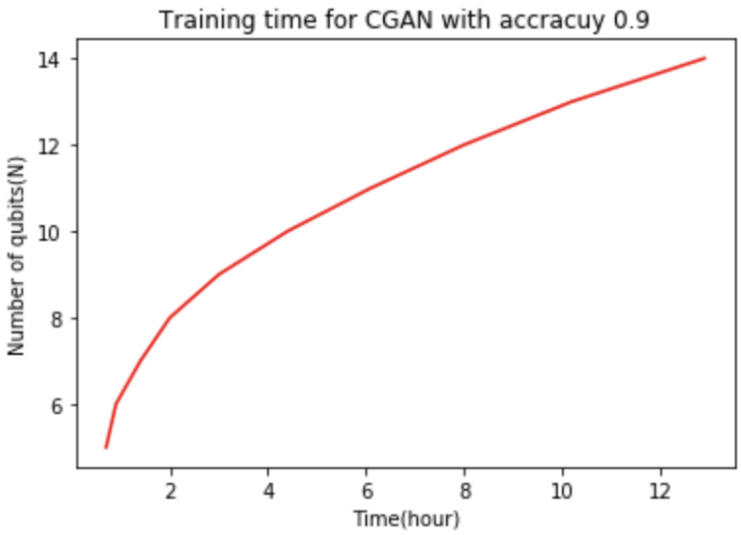

Figure 5.2: Training time vs Size of system

prediction based on classical shadow, the limitation of k-means clustering will be that before the prediction we must need to know some prior knowledge for this case which is periodicity and sparsity.

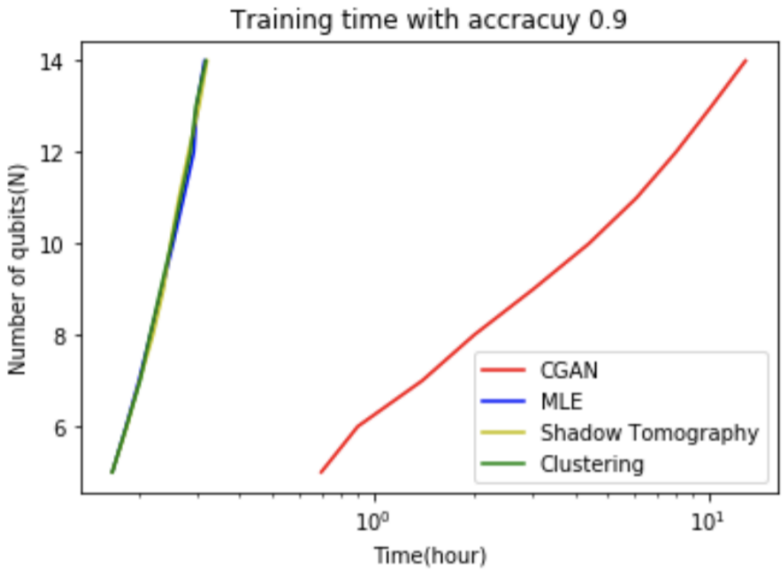

Figure 5.3: Training time vs Size of system

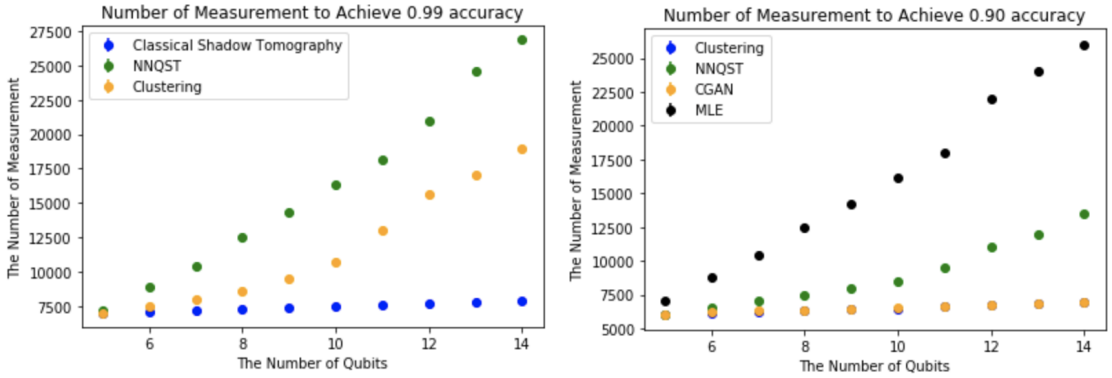

Figure 5.4: Comparison among each quantum tomography method

Chapter 6

Conclusion

We compared the performance of four different methods in predicting the result of period finding problem with error-included data. The analysis shows that given the prior knowledge about sparsity, the K-means clustering, classical shadow only require polynomial number of samples to achieve 0.99 accuracy. The CGAN method also need a few amount of data but it take more resources (time and samples) in training step. In the future, with the increasing size of quantum computer, combining the classical shadow tomography with other statistical method and generalizing the neural network method to higher dimension with few data will be two interesting ways to dig in.

References

[1] Shahnawaz Ahmed, Carlos Sánchez Muñoz, Franco Nori, and Anton Frisk Kockum. Classification and reconstruction of optical quantum states with deep neural networks. 12 2020.

[2] Ronen Barak and Yacob Ben-Aryeh. Quantum fast fourier transform and quantum computation by linear optics. *Journal of the Optical Society of America B*, 24(2):231, 2007.

[3] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[4] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, 1982.

[5] Min Xiao, Ling-An Wu, and H. J. Kimble. Precision measurement beyond the shot-noise limit. *Physical Review Letters*, 59(3):278–281, 1987.

[6] G. Y. Xiang, B. L. Higgins, D. W. Berry, H. M. Wiseman, and G. J. Pryde. Entanglement-enhanced measurement of a completely unknown optical phase. *Nature Photonics*, 5(1):43–47, 2010.

[7] M. W. Mitchell, J. S. Lundeen, and A. M. Steinberg. Super-resolving phase measurements with a multiphoton entangled state. *Nature*, 429(6988):161–164, 2004.

[8] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde. Entanglement-free heisenberg-limited phase estimation. *Nature*, 450(7168):393–396, 2007.

[9] Thibault Damour, Bala R. Iyer, and B. S. Sathyaprakash. Erratum: Comparison of search templates for gravitational waves from binary inspiral: 3.5pn update [phys. rev. d66, 027502 (2002)]. *Physical Review D*, 72(2), 2005.

[10] U. Fano. Description of states in quantum mechanics by density matrix and operator techniques. *Reviews of Modern Physics*, 29(1):74–93, 1957.

[11] Robin Blume-Kohout. Hedged maximum likelihood quantum state estimation. *Physical Review Letters*, 105(20), 2010.

[12] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical Review Letters*, 105(15), 2010.

[13] Wei-Tao Liu, Ting Zhang, Ji-Ying Liu, Ping-Xing Chen, and Jian-Min Yuan. Experimental quantum state tomography via compressed sampling. *Physical Review Letters*, 108(17), 2012.

[14] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2010.

[15] E. Knill, R. Laflamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409(6816):46–52, 2001.

[16] R. O'Donnell and John Wright. Efficient quantum tomography. *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 2016.

[17] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, page 1–1, 2017.

[18] Scott Aaronson. Shadow tomography of quantum states. *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, 2018.

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[20] Xudong Mao and Qing Li. Generative adversarial networks (gans). *Generative Adversarial Networks for Image Generation*, page 1–7, 2020.

[21] Iris Agresti, Niko Viggianiello, Fulvio Flamini, Nicolò Spagnolo, Andrea Crespi, Roberto Osellame, Nathan Wiebe, and Fabio Sciarrino. Pattern recognition techniques for boson sampling validation. *Physical Review X*, 9(1), Jan 2019.